

Genetic diversity of mango accessions (*Mangifera indica*) using new microsatellite markers and morphological descriptors

Elaini Oliveira dos Santos Alves¹, Francisco Pinheiro Lima Neto², Carlos Antônio Fernandes Santos², Ierla Carla Nunes dos Santos Ribeiro², Cláudio Antônio Ferreira de Melo¹, Ioná Santos Araújo Holanda³, Anete Pereira de Souza^{4,5}, Ronan Xavier Corrêa^{1*}

¹UESC, Universidade Estadual de Santa Cruz, Departamento de Ciências Biológicas, Centro de Biotecnologia e Genética, Rodovia Jorge Amado, km 16, CEP 45662-900 – Ilhéus, BA, Brazil

²Embrapa Semiárido, Empresa Brasileira de Pesquisa Agropecuária. BR 428 KM 152. Zona Rural 56300970 - Petrolina, PE - Brazil - Caixa-postal: 23

³UFERSA, Universidade Federal Rural do Semiárido, Av. Francisco Mota, Bairro Costa e Silva, CEP 59.625-900 -Mossoró, RN Brazil

⁴UNICAMP, Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Biologia Vegetal CEP 13083-862, Campinas, São Paulo, Brazil

⁵UNICAMP, Universidade Estadual de Campinas, Centro de Biologia Molecular e Engenharia Genética, CEP 13083-875, Campinas, SP, Brazil

*Corresponding author: ronanxc@uesc.br

Abstract

Genetic diversity estimates based on morphological and molecular data can provide different information on the relationship between cultivars of a species. This study aimed to develop new microsatellite markers as additional tools in genetic studies on mangoes (*Mangifera indica* L.), and to analyze the genetic variability of 20 mango cultivars based on morphological descriptors and microsatellite markers. We aimed to better understand the cultivars enhanced breeding histories and to support crossbreeding planning. Positive clones were selected from a DNA library enriched for microsatellite regions for sequencing and primer design. Four plants of each of the 20 accessions were used for observations, based on 48 morphological descriptors. Twenty accessions were analyzed using 27 microsatellite markers, of which 16 were developed during this study. The clusters, based on the morphological descriptors by Ward-MLM strategy and the microsatellite markers, suggested that Brazilian mango cultivars have extensive genetic diversity and are related to cultivars with different provenances, demonstrating their different enhanced breeding histories.

Keywords: genetic breeding, germplasm, molecular marker, SSR, Ward-MLM.

Introduction

The mango (*M. indica*) is the most widespread of the 69 species of the genus *Mangifera*, from the Anacardiaceae family (Kostermans and Bompard, 1993). The fruit's suitability for cultivation is well documented and over the centuries, its commercialization has grown extensively. The Portuguese first introduced the mango tree to Brazil, bringing polyembryonic cultivars to the Brazilian state of Bahia from Southeast Asia in the 16th century; these resulted in fibrous fruits that were also called the 'Filipino breed'. These polyembryonic types were subsequently grown in Brazil, which resulted in many different regional mango trees, which came from the planting of seeds, such as the 'Espada' and the 'Rosa'. Monoembryonic cultivars were introduced from the middle of the 20th century, provided by enhanced genetic breeding in North America, such as 'Tommy Atkins', 'Keitt', 'Palmer', 'Haden' and 'Van Dyke' (Pinto et al., 2002; Soares, 2000). The introduction of these cultivars, coupled with the adoption of irrigation in Northeastern Brazil, caused a substantial change in the cultivation of this fruit tree, characterized by much improved yields.

The search for promising new cultivars requires knowledge and understanding of their genetic diversity, because genetic enhancement can lead to the genetic base becoming less rich. Thus, knowledge of genetic diversity could be used to identify hybrid combinations with a greater heterotic effect, thereby increasing the possibility of finding superior genotypes in the segregant generations (Cruz et al., 2004) and decreasing inbreeding depression, which is a serious problem for *M. indica*. To identify cultivars and studying genetic diversity, the characteristics of the fruit, leaves, panicle and others parts have been used successfully as conventional genetic markers in *M. indica*. However, their morphological characteristics have low levels of polymorphism and are sensitive to environmental change. Thus, many genetic diversity studies based on morphological characterization have not reflected the individuals' actual genetic diversity. Using molecular markers, all DNA information can be targeted in the evaluation, depending on the nature of the marker used. Molecular markers are considered neutral, i.e. they are not influenced by environmental conditions. In addition, if they are not linked to a gene of agronomic

interest, they will be under little or no selective pressure in the culture's enhanced breeding process. Molecular markers have therefore been used regularly for diversity studies in mangoes (Díaz-Matallana et al., 2009; Ravishankar et al., 2004; Kumar et al., 2001; Santos et al., 2008).

However, many morphological characteristics are influenced by selection through many generations and represent the culture's enhanced breeding history. Therefore, morphological characterization is essential, especially in enhanced breeding programs. Thus, genetic diversity estimates that are based on morphological and molecular data provide different relational information among cultivars, which are important for choosing genitors for hybridizations and obtaining information regarding the diversity of a germplasm bank.

The aim of this study was to develop and characterize new microsatellite markers as an additional tool for genetic studies in mangoes. Additionally, the genetic variability of 20 mango accessions, mostly Brazilian cultivars, were analyzed using the Ward-modified location model (Ward-MLM) strategy to evaluate morphological descriptors and microsatellite markers for other clustering methodology. The genetic diversities of certain Brazilian cultivars and their relationships with foreign cultivars were analyzed using this approach, with the aim of better understanding these cultivars' enhanced breeding histories and supporting crossbreeding planning to produce superior cultivars.

Results

New SSR markers for mango

Microsatellite loci were obtained from the 48 positive clones, allowing us to design 32 primer pairs based on the flanking sequences of the microsatellite loci. Among these, 16 primer pairs amplified PCR products typical of microsatellite markers (Table 1), and were used to amplify DNA from 20 accessions of mango (see raw data in Supplementary Table 2). We show the complete sequences of genomic clones containing microsatellites (GenBank KR998053-KR998068). The mean number of alleles per locus was four (ranging from two to six). The mean values of observed (H_o) and expected (H_e) heterozygosity were 0.438 (0 to 1) and 0.625 (0.405 to 0.794), respectively. The average PIC value was 0.551, suggesting that the analyzed microsatellite loci were moderately informative (Hildebrand et al., 1992).

Genetic diversity in mango cultivars based on SSR markers

The data from the microsatellite markers developed in this study were used, in conjunction with data obtained by Ribeiro et al. (2012) (27 microsatellites in total), to cluster the 20 mango accessions for this analysis. The dendrogram showed five clusters (Fig. 1), which consisted of two large clusters comprising eight accessions each, two clusters comprising one accession in each and the last group containing two accessions. The smallest similarity coefficient among the accessions was found between the accessions 'Princesa' and 'Manzanillo' (0.493), while the largest similarity coefficient was observed between 'Keitt' and 'Recife' (0.839).

Most of the accessions used in this study were Brazilian. In this cluster, Brazilian cultivars 'Recife', 'Rosari', 'Princesa', 'Espada Itaparica', 'Comprida Roxa', 'Juazeiro VI' and 'Ametista' were related to the American cultivar 'Keitt'. Other Brazilian cultivars ('Ipuçaba', 'Foice', 'Juazeiro III' and 'Surpresa') were clustered with the Australian cultivar

'Kensington', as well as with American cultivars 'Edward' and 'Haden'. Brazilian cultivars 'Pingo de Ouro DPV' and 'Pêssego DPV' were clustered in isolation from the other groups, while the American cultivar 'Palmer' was clustered with the Mexican cultivar 'Manzanillo'.

Genetic diversity in mango cultivars based on morphological data

The calculated similarity coefficients among the 20 mango accessions, based on qualitative morphological data, ranged from 0.235 (between cultivars; 'Espada Itaparica' and 'Manzanillo') to 0.686 (between cultivars; 'Juazeiro III' and 'Keitt', and between 'Haden' and 'Edward') (results not showed). In the dendrogram generated by the UPGMA method and based on qualitative data, the 20 studied accessions were separated into four clusters (Fig. 2A). The cultivar 'Surpresa' formed a cluster separate from the other accessions, but was more related to the accessions of the cluster comprising 'Edward', 'Haden', 'Palmer', 'Juazeiro III', 'Keitt' and 'Foice'. This second cluster contained all the American cultivars used in this study. The other two clusters were segregated from the two groups already mentioned. Brazilian cultivars 'Juazeiro II', 'Rosari', 'Recife', 'Comprida Roxa', 'Pêssego DPV' and 'Ametista' were more related to the Australian cultivar 'Kensington' and to the Mexican cultivar 'Manzanillo', while the fourth cluster only comprised Brazilian cultivars. However, this last group is more closely related to the cluster of the Australian and Mexican cultivars than to the American ones.

The calculated genetic distances, based on quantitative data, ranged from 0.121 (between 'Keitt' and 'Haden') to 3.012 (between 'Princesa' and 'Juazeiro III') (results not showed). The dendrogram, based on genetic distances, showed four clusters (Fig. 2B). One of them could be subdivided: one subcluster comprised the four American cultivars ('Palmer', 'Edward', 'Keitt' and 'Haden') and the Brazilian cultivar 'Surpresa', and other subgroup comprised some Brazilian cultivars ('Foice', 'Rosari', 'Recife' and 'Espada Itaparica'), the Australian 'Kensington' and the Mexican 'Manzanillo'. A second large cluster could also be divided into two, each containing only Brazilian cultivars. One cluster contained the cultivar 'Comprida Roxa', which may be more related to the cluster of Brazilian cultivars. The cultivar 'Princesa' comprised the fourth cluster, which was the most isolated from the others.

The Ward-MLM procedure determined that the optimal number of clusters was nine (Fig. 3A). The clusters were: 1- 'Edward', 'Haden' and 'Palmer'; 2- 'Kensington', 'Juazeiro II', 'Juazeiro VI' and 'Pingo de Ouro DPV'; 3- 'Recife', 'Manzanillo', 'Pêssego DPV', 'Ametista' and 'Comprida Roxa'; 4- 'Keitt' and 'Princesa'; 5- 'Ipuçaba'; 6- 'Rosary' and 'Foice'; 7- 'Surpresa'; 8- 'Juazeiro III'; and 9- 'Espada Itaparica'.

The cultivars in cluster 1 were the American cultivars; cluster 2 comprised Brazilian cultivars and one Australian; cluster 3 was mostly Brazilian and one Mexican; cluster 4 comprised one American and one Brazilian, and the remaining clusters comprised Brazilian cultivars. This large number of clusters suggested significant diversity among the analyzed accessions.

The first two canonical variables explained 92.27% of the variability among the nine clusters (Fig. 3B), indicating that the graphic representation of the first two canonical variables ('latex quality' and 'skin weight') was suitable for visualizing the genetic relationship among the clusters, as well as among the accessions within the same group. "Latex Quality"

Table 1. Characteristics of 16 microsatellite loci isolated from mango DNA samples.

SSR Locus	Primer sequences (5'-3')	Allele size (bp)	Repeat motif	NA	Ho	He	PIC
MiL01	F: TCGCTTATGGCTCCAAGTTT R: GGCAATGGTCCTGATGAAGT	181-211	(TG)6	4	0.100	0.544	0.489
MiL02	F: GAAGTGGCAGCTGTATGTGC R: CCTTTTCTGGGTCATGGAGA	330-340	(TTA)3	2	1.000	0.513	0.375
MiL04	F: TGAGCCCAAATTGAGATTGTC R: CTTGCTGTTCTTGCCATCAA	204-210	(TG)10	4	0.750	0.673	0.598
MiL07	F: TCCCACTTGGATAGCATTGA R: ATTTGGGTGCTATTCTTGCC	290-318	(TG)12	5	0.400	0.765	0.704
MiL08	F: TCGGTTTTCGATTGATAACCTC R: AAAGCATCGGTAGTCGGTTG	308-330	(ATT)3	4	0.050	0.747	0.679
MiL09	F: CGAACGTACCCCATCAAAAAG R: TGTGACAGGTCTGTCTGGTG	122-126	(ATGAG)2	3	0.000	0.456	0.381
MiL13	F: GGTCACACACAGAAGGGGTT R: GCCAATTCTCGCATAACCT	172-178	(TAT)3	3	0.500	0.549	0.436
MiL14	F: TGGTTTACTGTGCACATGCC R: CTAGCCCCGAACATGAAGAG	238-242	(AC)15	3	0.200	0.621	0.527
MiL15	F: TATCAGCCAATTTTGCCCTT R: GAGGTGGACCATAGGGGTTT	312-320	(TTTGT)2	3	0.400	0.405	0.347
MiL16	F: ATGAGCCGATTGGGCTATTA R: TGATCAAATGTGGCTTGCTC	243-253	(TG)6C(GT)6	5	0.600	0.673	0.605
MiL20	F: GGATTGACAAGGAGGGGAAT R: ATTTGGCGTTTTGAAACCTG	208-216	(GA)4...(AT)6(GT)9	4	0.450	0.609	0.541
MiL21	F: AACCGGAGATGCTGAAATTG R: ATTGCAGGAACCATCCTTCA	186-210	(TAT)4	5	0.600	0.583	0.516
MiL23	F: CCTCCAATCTCCTTCTTCC R: AGCCATCTTTTTCTCCTCGTT	294-300	(ATATC)3	3	0.000	0.518	0.442
MiL27	F: AAGACATCATGGCCACTTGAC R: CGAAGACCATGGTGGATTA	333-337	(TG)9	5	0.400	0.794	0.736
MiL29	F: AATGACAATGGGGGTGAAAA R: GTTCGGAGAAAAGTGTGGGA	153-159	(ATTCCC)2	5	0.850	0.786	0.729
MiL37	F: TTGGGTATCCTTTGGAGTGC R: CAGCCTGAAAATGCAAGAGA	331-345	(CA)11...(AT)3...(GAA)2	6	0.700	0.763	0.703

Note: Ho= observed heterozygosity; He=expected heterozygosity; Na=number of alleles; PIC= polymorphism information content.

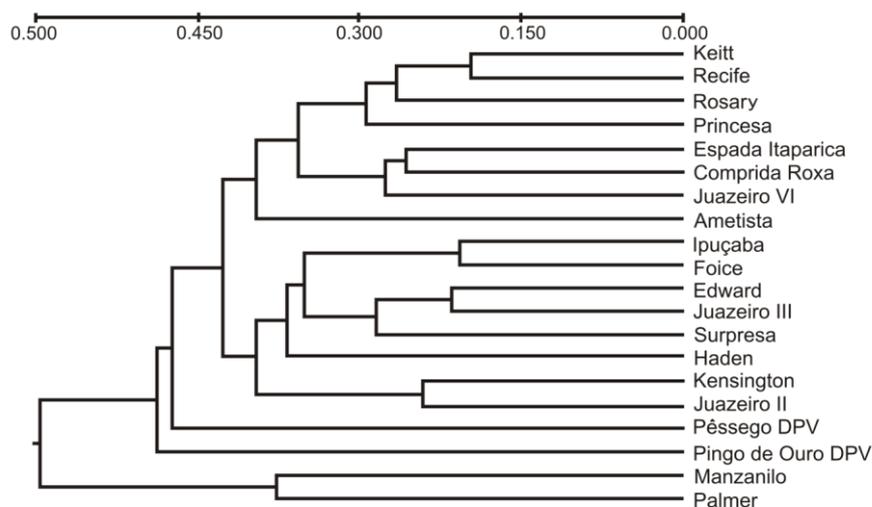


Fig 1. Clustering of 20 mango cultivars, generated by the UPGMA clustering method, using a similarity matrix generated by Nei's standard genetic distance based on data from 27 microsatellite markers.

accounted for most of the variation with the first canonical variable, followed by "Fruit Length" and "Skin Weight", with values of 0.679, 0.569 and 0.423, respectively. "Fruit Length" and "Skin Weight" are targeted more frequently in the selection of superior genotypes (Table 2). The majority of American cultivars used in this study clustered together. The criterion used for cluster separation, while considering the canonical variables, was associated with the enhanced breeding history of the accessions.

Discussion

Analysis using molecular markers has generated useful information regarding the genetic diversity in agricultural species. According to Dillon et al. (2014), high quality genetic analysis of species such as *M. indica* requires a large number of informative polymorphic markers, which have been lacking in *M. indica* to date.

The 16 new, specific, microsatellite markers for *M. indica* developed during this study were moderately informative. Adopting the considerations of Hildebrand et al. (1992), loci MiL07, MiL27, MiL29 and MiL37 could be classified as highly informative, because they reached PIC values greater than 0.7; however, loci MiL02, MiL09 and MiL15 could be considered as having low informational value. The remaining nine loci are moderately informative. Schnell et al. (2005), Ravishankar et al. (2004) and Chiang et al. (2012), who developed microsatellite markers for *M. indica*, also obtained mean PIC values that were considered moderately informative. According to Hildebrand et al. (1992), the more repetitions a locus has and the less the sequence is repeated, the greater the tendency for the locus to be highly informative. The results found in this study corroborated this conclusion for the majority of the loci. Among the 16 SSR markers, eight have microsatellites composed of dinucleotides, of which three were highly informative and the other five were moderately informative. These markers increased the number of available markers for the genetic analysis of mango tree cultures.

The accessions that were used in the characterization of these loci were previously analyzed using eleven microsatellite markers (Ribeiro et al., 2012). Most of these accessions are considered Brazilian cultivars and little is known regarding the history of their origin. One cluster analysis was jointly performed, using the data previously

obtained by Ribeiro et al. (2012), together with the 16 loci analyzed in this study. This analysis showed that the Brazilian cultivar 'Recife' has a very close genetic relationship with the American cultivar 'Keitt'. The Brazilian cultivars 'Rosary' and 'Princesa' also have a close relationship with 'Keitt'. The Brazilian cultivars 'Juazeiro III' and 'Surpresa' proved to be genetically close to the American cultivar 'Edward', while 'Juazeiro II' is related to the Australian cultivar 'Kensington'. The American cultivar 'Palmer' clustered with the Mexican cultivar 'Manzanillo'.

While studying the similarity of accessions belonging to the same germplasm bank that were used in this study, and based on amplified fragment length polymorphism (AFLP) markers, Santos et al. (2008) suggested that the accession identified as 'Palmer' might not be a clone of the American 'Palmer' or that it had been erroneously identified, because they observed that 'Palmer' clustered with polyembryonic cultivars. In the present study, this cultivar clustered with one that is considered predominantly monoembryonic. It is worth highlighting that this 'embryony' feature probably comes from a monogenic heritage (Arnon et al., 1998), which makes it difficult to identify a strong association of this character with the formation of clusters in genetic diversity studies, when based on molecular markers such as microsatellites, which access random DNA regions. Nevertheless, this feature may be related to other characteristics, which would justify an embryonic type association as a criterion of weight during clustering. Ravishankar et al. (2004), while analyzing mango cultivars with random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers, which also access random DNA regions, observed segregation of two clusters associated with the embryonic type of the cultivars and inferred that the embryonic types of mango cultivars have a different genetic basis. In this study, no association of embryonic type with the clusters' formation was verified.

In the cluster based on morphological descriptors, both based on qualitative and quantitative data separately, as well as jointly, the cultivar 'Palmer' was clustered with other American cultivars, which are also monoembryonic, reflecting both an association with the genetic base associated with embryony, as well as with the enhanced breeding of American cultivars.

Table 2. Averages of 10 quantitative variables for each of the clusters formed by the Ward-MLM method and the first two canonical (CAN) variables.

Variables	Groups									CAN	
	1 (3)	2 (4)	3 (5)	4 (2)	5 (1)	6 (2)	7(1)	8 (1)	9 (1)	CAN1	CAN2
D46	58.57	30.13	42.52	54.00	36.50	40.10	23.90	23.80	41.20	0.419	-0.569
D32	207.63	102.25	102.84	101.95	79.50	86.70	94.10	183.90	72.50	0.320	-0.061
D34	1.07	0.88	2.00	0.95	0.90	0.85	0.40	0.60	0.80	-0.211	-0.169
D35	104.00	32.38	55.82	77.75	26.50	58.95	44.80	21.90	45.30	0.423	-0.515
D29	0.73	0.43	0.22	0.60	1.20	1.15	0.50	1.00	0.60	0.679	0.505
D3	3.03	3.30	3.44	2.90	3.00	3.75	3.00	4.70	4.70	-0.275	0.737
D11	28.79	23.38	27.08	33.45	28.00	28.20	29.80	26.10	35.60	0.386	-0.183
D12	18.30	18.00	21.90	21.20	12.60	19.65	16.10	17.80	23.50	-0.264	-0.153
D14	13.13	7.38	9.38	11.80	8.40	11.60	10.70	8.40	9.60	0.569	-0.311
D15	8.70	7.63	8.70	9.00	7.60	9.70	7.80	7.00	6.50	0.130	-0.286

(D46) Seed Weight; (D32) Lenticel Density; (D34) Skin Thickness; (D35) Skin Weight; (D29) Latex Quantity; (D3) Petiole Length; (D11) Inflorescence Length; (D12) Inflorescence Width; (D14) Fruit Length; (D15) Fruit Width.

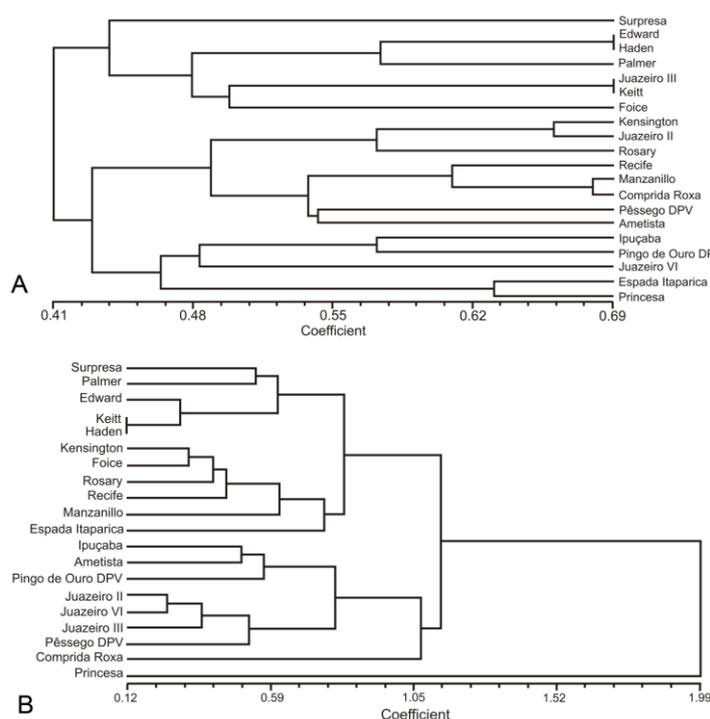


Fig 2. Clustering of 20 mango cultivars. The clustering was generated by the UPGMA method from two different data sets: A) a similarity matrix generated using Nei's Standard Genetic Distance based on qualitative data ($r=0.62$); B) a dissimilarity matrix generated using the mean Euclidean distance based on quantitative data ($r=0.78$).

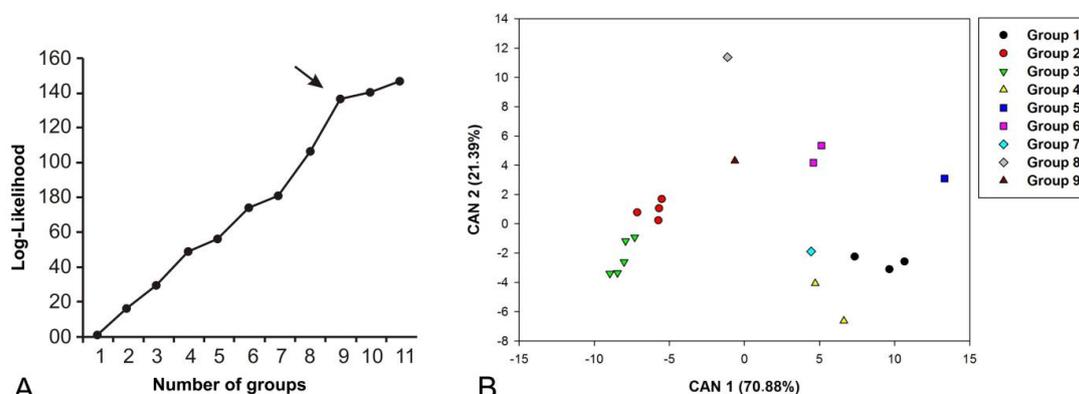


Fig 3. Joint multivariate analysis of quantitative and qualitative data by the Ward-MLM strategy. (A) Logarithmic probability function (Log-Likelihood) with the formation of nine clusters (as indicated by arrow). (B) The first two canonical variables for the nine clusters explaining 92.27% of the variation (colours indicate the group number).

The largest similarity coefficient observed (between 'Edward' and 'Haden') could, in part, be explained by the fact that they are American cultivars, produced by enhanced genetic breeding. Cultivars; 'Juazeiro III' and 'Keitt' also showed a similarly high coefficient, suggesting that there is a relationship between them, which may explain the origin of the Brazilian cultivar.

The clustering based on the morphological descriptors corroborated some observations based on the molecular data. The Brazilian cultivar 'Juazeiro II' has a strong genetic relationship with the Australian cultivar 'Kensington'. This relationship was not observed during clustering that only considered quantitative data, where the cultivar 'Juazeiro II' was clustered closer to other Brazilian cultivars. Note that most of the quantitative features used in this study had been used in the enhanced breeding process, in addition to the greater influence of the environmental conditions, because these features are believed to be controlled by many genes. Despite the quantitative characteristics being a source of great variation, those that are most frequently targeted for selection tended to have a tapered genetic variability as a result of the enhanced breeding process. Therefore, when these characteristics are used in genetic diversity studies, they demonstrate a greater similarity among enhanced cultivars, even with different genetic bases.

Despite being a polyembryonic cultivar, the Brazilian cultivar 'Surpresa' was clustered with the American monoembryonic cultivars in three clusters (based on molecular and morphological data). However, in the cluster based on the quantitative data (Fig. 2B), its relationship with one American cultivar was closer than in the other cases, suggesting that this relationship may be more related to the enhanced breeding history than to the genetic base. During the joint analysis, this cultivar was separated by itself in an isolated group.

The Brazilian cultivar 'Rosary' was clustered to the Australian cultivar 'Kensington' in the clusters based on separately performed morphological data, but showed no significant relationship in the clusters based on molecular data or when a joint cluster analysis of the morphological data was made. In addition, the Brazilian cultivar 'Recife' was related to the Mexican 'Manzanillo' in clusters based on morphological data, both when used separately and combined; however, they were not related in the clusters based on molecular markers.

In the dendrogram generated from genetic distances based on quantitative morphological descriptors, most Brazilian cultivars were allocated in separate clusters from the foreign cultivars. This was also observed when the clustering was performed based on qualitative and quantitative characteristics jointly.

Materials and Methods

Plant materials

The Banco Ativo de Germoplasma de Mangueira da Embrapa Semiárido (a mango Active Germplasm Bank—AGB—in northeast Brazil) provided 20 accessions, most of which were Brazilian cultivars; 14 from Brazil, 4 from United States, one from Australia, and one from Mexico (Supplementary table 1). The city where the AGB is located (Juazeiro, coordinates 09°24'S, 40° 26'W) has a semi-arid climate, with 541.6 mm mean annual rainfall and 58.8% relative humidity. The city's average annual temperature is 26.8°C and it has a Vertisol type soil. This AGB maintains 162 accessions, 103 of which were characterized by Ribeiro et al. (2013). Each accession

from this AGB comprised four plants planted with a 10-m × 10-m spacing.

Development of new SSR markers

A DNA library enriched for microsatellite regions was developed according to Billote et al. (1999). The plant genomic DNA from the cultivar 'Tommy Atkins' was isolated from foliar sample tissues using the Cetyl trimethylammonium bromide (CTAB) method (Doyle and Doyle 1990). The DNA was digested with restriction enzyme *AfaI* and its fragments were linked to adapters Rsa 21 (5' CTCTTGCTTACGCGTGGACTA 3') and Rsa 25 (5' TAGTCCACGCGTAAGCAAGAGCACA 3') and then preamplified using complementary primers to the adapters. The amplified products were incubated with (CT)₈ and (GT)₈ biotinylated microsatellite probes for a period of 10 minutes to achieve hybridization. Subsequently, DNA fragments were captured using magnetic spheres that were coated with streptavidin (Promega Corporation, Fitchburg, WI, USA). These selected fragments were amplified using PCR, cloned into vector *pGEM@-T* and transformed into *Escherichia coli* XL1-BLUE in order to obtain a DNA library enriched for sequences of microsatellites. Forty-eight positive clones from this DNA library were selected for sequencing. The sequences were edited using BioEdit software - version 7.2.3 (Hall, 1999) to assemble contigs. The Chromas software (version 2.4.1) was used to eliminate the adapter sequences. An SSRLocator program was used to identify the microsatellite regions and the Primer3 software (Rozen and Skaletsky, 2000) was used to design the primers.

PCR was performed to amplify DNA from the 20 mango accessions using the new primers. The reactions (total volume of 10 µL) included 20 ng of DNA, 0.8 µM of each primer, 1 µL of 10 × PCR buffer, 4 mM of dNTP mix and 2 mM of MgCl₂. The following PCR program was used for all loci: 94°C for 2 min; 35 cycles at 94°C for 30 s, 54°C for 45 s, 72°C for 1 min; and a final extension of 6 min at 72°C. The amplification products were subjected to vertical electrophoresis in polyacrylamide denaturing gels (6%), which were stained with silver nitrate (Creste et al., 2001). The alleles' sizes were estimated by comparing them a 10-kpb DNA ladder (Invitrogen, Carlsbad, CA, USA).

Morphological Characterization

Four plants from each accession were used for morphological observation, based on 48 descriptors, comprising 10 quantitative descriptors and 38 qualitative descriptors. These descriptors are suitable for mango germplasm evaluation as previously demonstrated (Ribeiro et al., 2013).

Statistical analyses

The number of alleles per locus, the observed heterozygosity, the expected heterozygosity and the polymorphic information content (PIC) were calculated for the new microsatellite loci. Cervus software was used to characterize the new microsatellite loci and TFPGA software was used to cluster the accessions. We used the results of genotyping with the new microsatellite markers in conjunction with previous results of genotyping with 11 loci (Ribeiro et al., 2012) to calculate the genetic distances among cultivars, using Nei's genetic distances, and to perform the clustering analysis using the unweighted pair group method with arithmetic mean (UPGMA) method.

Data averages obtained using morphological characterization were used for multivariate analysis, which was performed using the UPGMA method based on the mean Euclidean distance for the quantitative data and Nei's standard genetic distance for the qualitative data. Mantel's test for cophenetic correlations was used to evaluate the clustering method's efficiency. Qualitative and quantitative morphological variables were jointly analyzed through the Ward-MLM procedure (Franco et al., 1998), using pseudo-T₂ and pseudo-F statistics to define the optimal number of clusters. The Ward grouping method was used, while considering the joint matrix obtained from the Gower algorithm (Gower, 1971). Multivariate statistical analyses were performed using the NTSYS software for the UPGMA method, while SAS software was used for the Ward-MLM procedure.

Conclusion

Estimating genetic variability is useful for improving crop performance, both for choosing genitors for cross breeding among divergent genotypes looking to exploit heterosis and to avoid genetic depression caused by inbreeding. The microsatellite markers that were developed in this study could be used for characterizing molecular genetic diversity in a larger number of mango germplasm accessions, while seeking greater understanding regarding the genetic basis of non-Brazilian plant varieties, such as from other provenances. According to morphological descriptors and molecular markers, it is possible to infer that the Brazilian mango cultivars are significantly genetically diverse when related to cultivars from different provenances. This relationship makes improved crossbreeding planning possible with a view to developing superior cultivars. Clustering based on data from morphological descriptors made it possible to observe a relationship with the enhanced breeding history of some accessions. By contrast clustering based on molecular data did not reflect that history because the molecular markers access random regions of the DNA, including those that have undergone selection pressure and those that have not.

Acknowledgments

Brazilian foundations that supported this research are "Fundação de Amparo à Pesquisa de São Paulo" (FAPESP, 2008/52197-4), "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES, PROCAD-NF2008, CAPES granted a scholarship to the first author), and "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq granted the research fellowships to APS and RXC).

Data archiving statement

The mango DNA sequences and SSR markers were submitted to GenBank under the accession numbers KR998053 - KR998068.

References

Arnon Y, Czosnek H, Gazit S (1998) Polyembryony in mango (*Mangifera indica* L.) is controlled by a single dominant gene. *Hortic Sci.* 33:1241-1242.

Billote N, Lagoda P, Risterucci AM, Baurens FC (1999) Microsatellite-enriched libraries: Applied methodology for the development of SSR markers in tropical crops. *Fruits.* 54:277-288.

Chiang YC, Tsai CM, Chen YK, Lee SR, Chen CH, Lin YS, Tsai CC (2012) Development and characterization of 20 new

polymorphic microsatellite markers from *Mangifera indica* (Anacardiaceae). *Ame J Bot.* 99:117-119 doi: 10.3732/ajb.1100443

Creste S, Tulmann-Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Report.* 19:299-306.

Cruz CD, Regazzi AJ, Carneiro PCS (2004) Modelos biométricos aplicados ao melhoramento genético. 3 ed., UFV, Viçosa.

Díaz-Matallana M, Schuler-García I, Ruiz-García M, Jaramillo EH (2009) Analysis of diversity among six populations of Colombian mango (*Mangifera indica* L. cvar, Hilacha) using RAPDs markers. *Electronic J Biotech.* 12:1-8.

Dillon NL, Innes DJ, Bally ISE, Wright CL, Devitt LC, Dietzgen RG (2014) Expressed Sequence Tag-Simple Sequence Repeat (EST-SSR) Marker Resources for Diversity Analysis of Mango (*Mangifera indica* L.). *Diversity.* 6:72-87. doi:10.3390/d6010072

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus.* 12:13-15.

Franco J, Crossa J, Villaseñor J, Taba S, Eberhart SA (1998) Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38:1688-1696.

Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics.* 27:857-874

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acid Symp Series.* 41:95-98.

Hildebrand CE, Torney DC, Wagner RP (1992). Informativeness of Polymorphic DNA markers. *Los Alamos Science.* 20: 100-102.

Kostermans AFGH, Bompard JM (1993) The Mangoes: Botany, Nomenclature, Horticulture and Utilization. Academic Press, London.

Kumar NVH, Narayanaswamy P, Prasad DT, Mukunda GK, Sondhur SN (2001) Estimation of genetic diversity of commercial mango (*Mangifera indica* L.) cultivars using RAPD markers. *J Horticult Science Biotechnol.* 76: 529-533.

Pinto ACQ, Souza VAB, Rossetto CJ, Ferreira FR, Costa JG (2002) Melhoramento Genético. In: Genú PJC, Pinto ACQ (Eds.). *A Cultura da Mangueira.* Embrapa Informação Tecnológica. p. 53-92.

Ravishankar KV, Chandrashekara P, Sreedhara SA, Dinesh MR, Anand L, Saiprasad GVS (2004) Diverse genetic bases of Indian polyembryonic and monoembryonic mango (*Mangifera indica* L) cultivars. *Current Sci.* 87:870-871.

Ribeiro ICNS, Lima-Neto FP, Santos CAF (2012) Allelic database and accession divergence of a Brazilian mango collection based on microsatellite markers. *Genet Mol Res.* 11:4564-4574.

Ribeiro ICNS, Santos CAF, Lima-Neto FP (2013) Morphological Characterization of Mango (*Mangifera indica*) Accessions Based on Brazilian Adapted descriptors. *J Agricult Sci Technol.* 3:798-806.

Rozen S, Skaletsky HJ (2000) Primer 3 on the WWW for general users and for biologist programmers. In S. Krawetz and S. Misener [eds.], *Bioinformatics methods and protocols: Methods in Molecular Biology,* Humana Press, Totowa, New Jersey. pp 365-386.

Santos CAF, Lima-Neto FP, Rodrigues MA, Costa JG (2008) Similaridade genética de acessos de mangueira de diferentes origens geográficas avaliadas por marcadores AFLP. *Rev Bras Frut.* 30:736-740.

Schnell RJ, Olano CT, Quintanilla WE, Meerow AW (2005) Isolation and characterization of 15 microsatellite loci from mango (*Mangifera indica* L.) and cross-species amplification in closely related taxa. *Mole Ecol Notes.* 5:625-627.

Soares NB (2000) Mangueira. In: Meletti LMM *Propagação de frutíferas tropicais.* Agropecuária, Guaíba, pp.178-187.